caBIG cancer Biomedical Informatics Grid

Cruising the Cancer Biomedical Informatics Grid

Peter A. Covitz, Ph.D.
NCI Center for Bioinformatics

February, 2003

NATIONAL CANCER INSTITUTE

caBIG: From Village to City

- Cars
- Fuel
- Driving School
- Services & Businesses
- Transit Routes
- Maps & Brochures
- Visitor Information

caBIG cancer Biomedical Informatics Grid

Cars, Fuel, and Services

- Cars fill up at stations and drive to useful destinations to buy products and services
- caBIG Applications will fill up on data and drive to analytical services
  - Data services will offer standardized representations of data
  - Analytic services will offer processing routines

Transit Routes

- Buses and cars cruise the street grid to get places
- caBIG Applications cruise caBIG to get data and other services
  - Service providers must support caBIG APIs and message standards
  - Applications will be caBIG API and message-aware

Driving School

- No one is born knowing how to drive!
- caBIG citizens will need training and tutorials to use caBIG APIs and applications

Maps

- You need maps to navigate a city
- You need documentation navigate caBIG
  - caBIG systems, code, applications and APIs will be fully documented
  - Getting it to work isn't good enough; others have to be able to use it

Getting Around:

Interoperability

Courtesy: Charlie Mead

Interoperability

- **in·ter·op·er·a·bil·i·ty**
  - ability of a system...to use the parts or equipment of another system

    Source: Merriam-Webster web site

- **interoperability**
  - ability of two or more systems or components to exchange information and to use the information that has been exchanged.

    Source: IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries, IEEE, 1990]

Syntactic interoperability

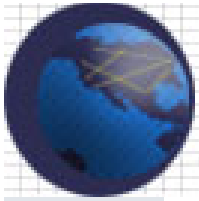
Semantic interoperability

Semantic Interoperability

Courtesy: Charlie Mead

Pillars of Interoperability

- Common models across all domains of interest
- Foundation of rigorously defined data types
- Methodology for interfacing with controlled vocabularies



caBIG cancer Biomedical Informatics Grid

Semantic Interoperability:

Common Models

NATIONAL CANCER INSTITUTE

What *is* a Model?

- Human-friendly picture of complexity
- Link to ‘lower-level’ models
  - Layering and segregation of complexity
  - Abstraction and separation of layers

Courtesy: Charlie Mead

Why build models?

- Models represent an important vehicle for reaching *consensus* about the architecture (structure and function) of a Problem and/or a Solution

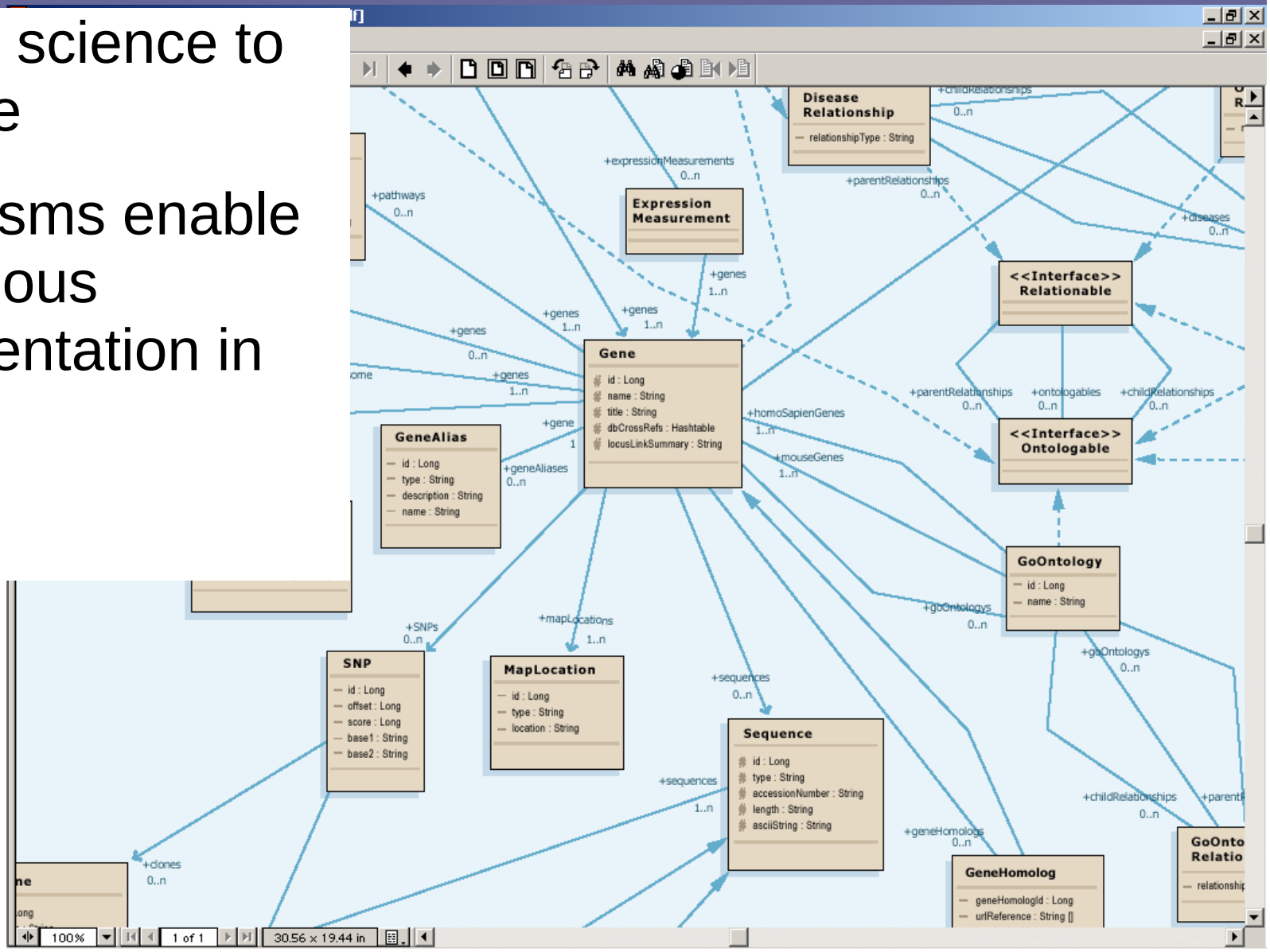
How to model in caBIG

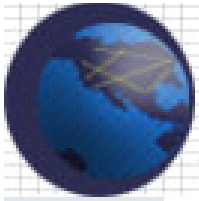
- Industry-standard best practices
- Collect use cases
  - If the system were already built, what would you use it to do, precisely?
- Define data classes and their attributes
- Identify data class relationships
- Construct the model in UML
- Review with stakeholders, refine
- Feed into software and database designs

NATIONAL CANCER INSTITUTE
caBIG cancer Biomedical Informatics Grid

Unified Modeling Language

- Bridges science to software
- Formalisms enable expeditious implementation in caBIG





caBIG cancer Biomedical Informatics Grid

Semantic Interoperability:

Common Data Elements

NATIONAL CANCER INSTITUTE

What is a CDE?

- Everything you need to describe and understand what a datum means
- Metadata about the individual questions and answers in a study
- A means towards semantic continuity and data comparability across studies over time

What CDEs provide to caBIG

- Solve problems of ambiguity
  - Precise definition of data types, all the way through to scientific meaning
- Save analysis time
  - Minimize need to reverse engineer meaning from data
- Enable comparability
  - Large, multi-institutional, multi-study data comparisons can provide more power

CDE development strategy for caBIG

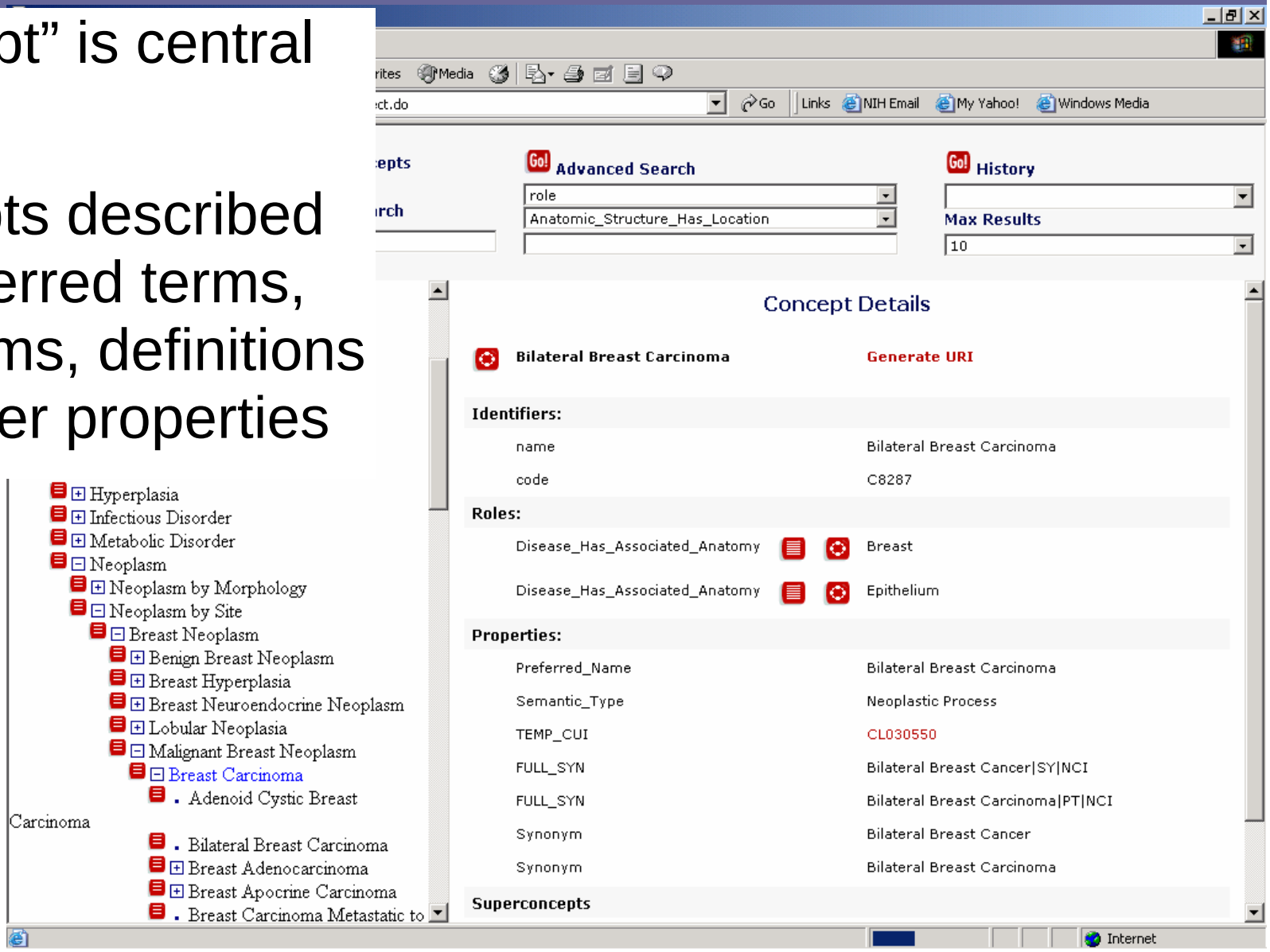
- Key Figures:
  - Investigator/study team
  - Domain experts
  - CDE Administrator
- Data elements identified as study protocols are created
  - Need-driven, not an abstract modeling exercise
- Existing CDEs re-used, new ones created as needed
  - External standards can also be represented as CDEs e.g. ICD-O-3
- Harmonization process to review CDEs and select preferred standards

Semantic Interoperability:

Common Vocabularies

What is a common vocabulary?

- "Concept" is central entity
- Concepts described by Preferred terms, synonyms, definitions and other properties



NATIONAL CANCER INSTITUTE

Why do we need Common Vocabularies in caBIG?

- CDEs and biomedical data classes are composite structures synthesized from multiple concepts
- The component concepts must be defined using common, reusable terminologies

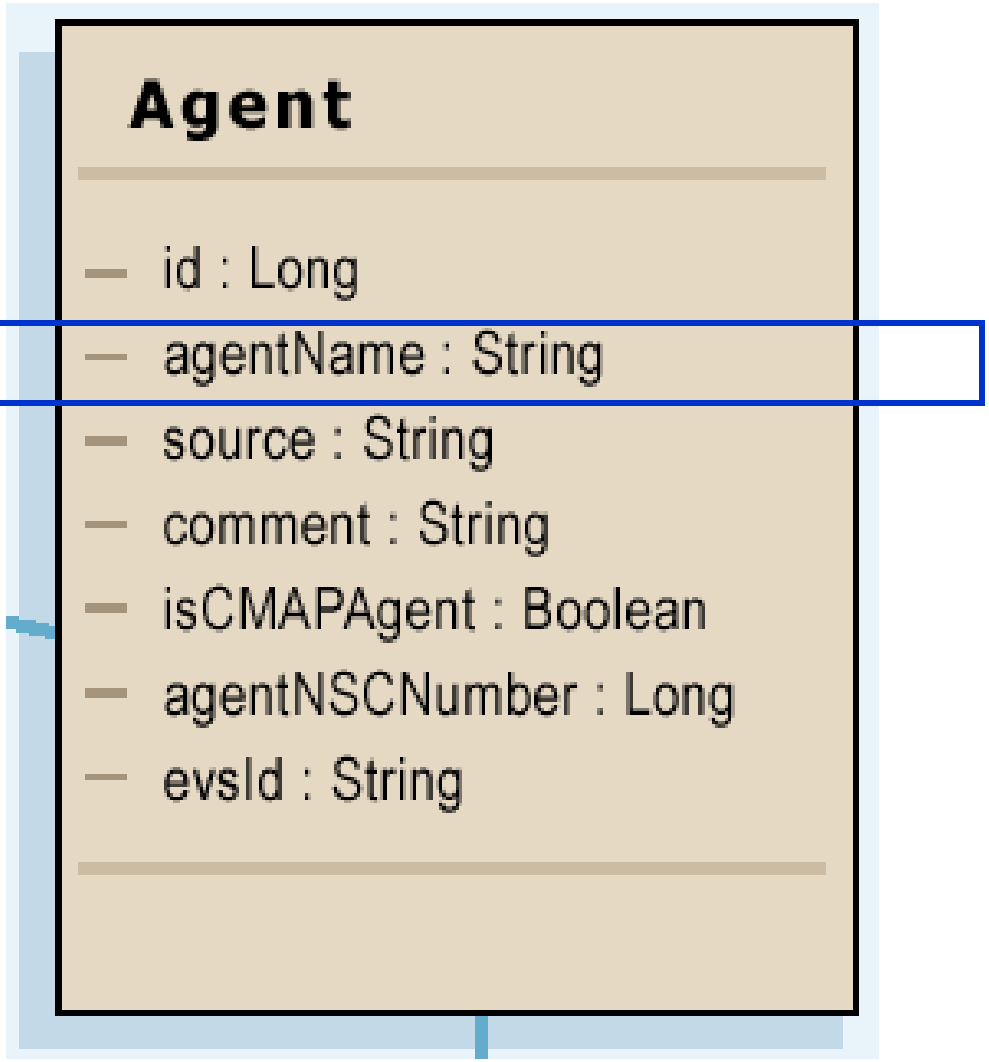
How are Vocabularies used in caBIG?

- Supply common terminology for CDE and UML data class development
- Provide data standards for valid values
- In a description logic framework, provide semantic linkages to related concepts

Semantic Interoperability:

Tying it all together

Common Model →



→ Common Data Element →

Value Domain Details

Public ID:	2018334
Preferred Name:	AGT_NAME
Long Name:	Agent Name
Definition:	the name of the agent or drug that has been administered to the patient.
Workflow Status:	RELEASED
Version:	1.0
Datatype:	CHARACTER
Unit of Measure:	
Display Format:	
Maximum Length:	100
Minimum Length:	
Decimal Place:	
High Value:	
Low Value:	
Value Domain Type:	Non Enumerated
Conceptual Domain Preferred Name:	TX
Conceptual Domain Context Name:	CTEP
Conceptual Domain Version:	1.0
Origin:	

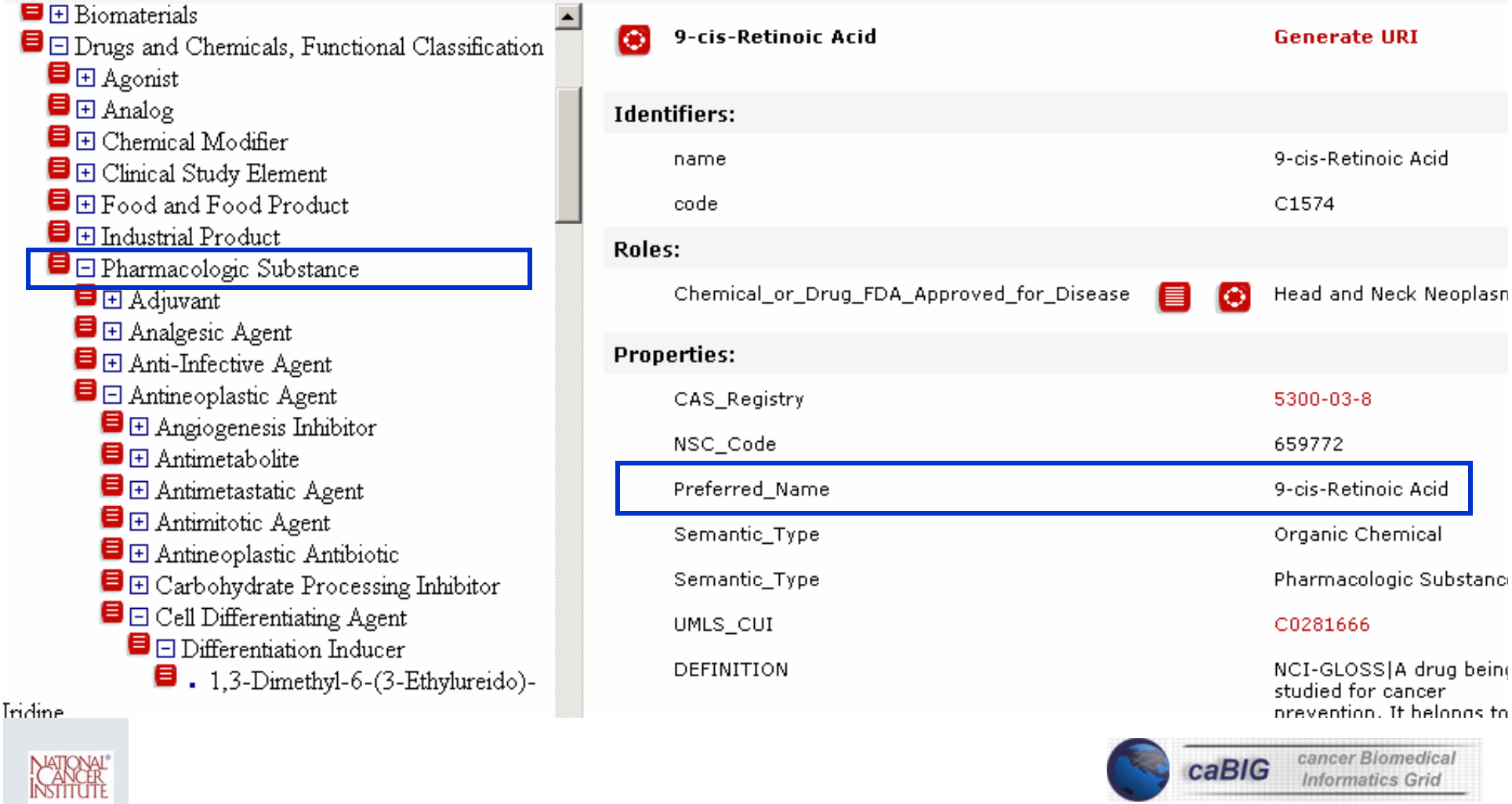
Permissible Values

This Value Domain is Non Enumerated

NATIONAL CANCER INSTITUTE

caBIG cancer Biomedical Informatics Grid

→ Common Vocabulary



NATIONAL CANCER INSTITUTE

caBIG cancer Biomedical Informatics Grid

Syntactic Interoperability:

Common APIs

Interchange Formats

Messaging Standards

Why common APIs, formats, and messages?

- Takes less time to learn how to access more kinds of data
- Dynamic access to data stores in real time
- System-to-system messaging enables sophisticated workflows with less human intervention

Accessible APIs for caBIG

- Aligned with common biomedical information models
  - APIs become natural extension of biomedical data domain
- Broad programming language support
  - No good if average bioinformatician can't use them!
- Extended according to a common paradigm
  - Developers only have to learn it once, then it is familiar

Interchange and message formats

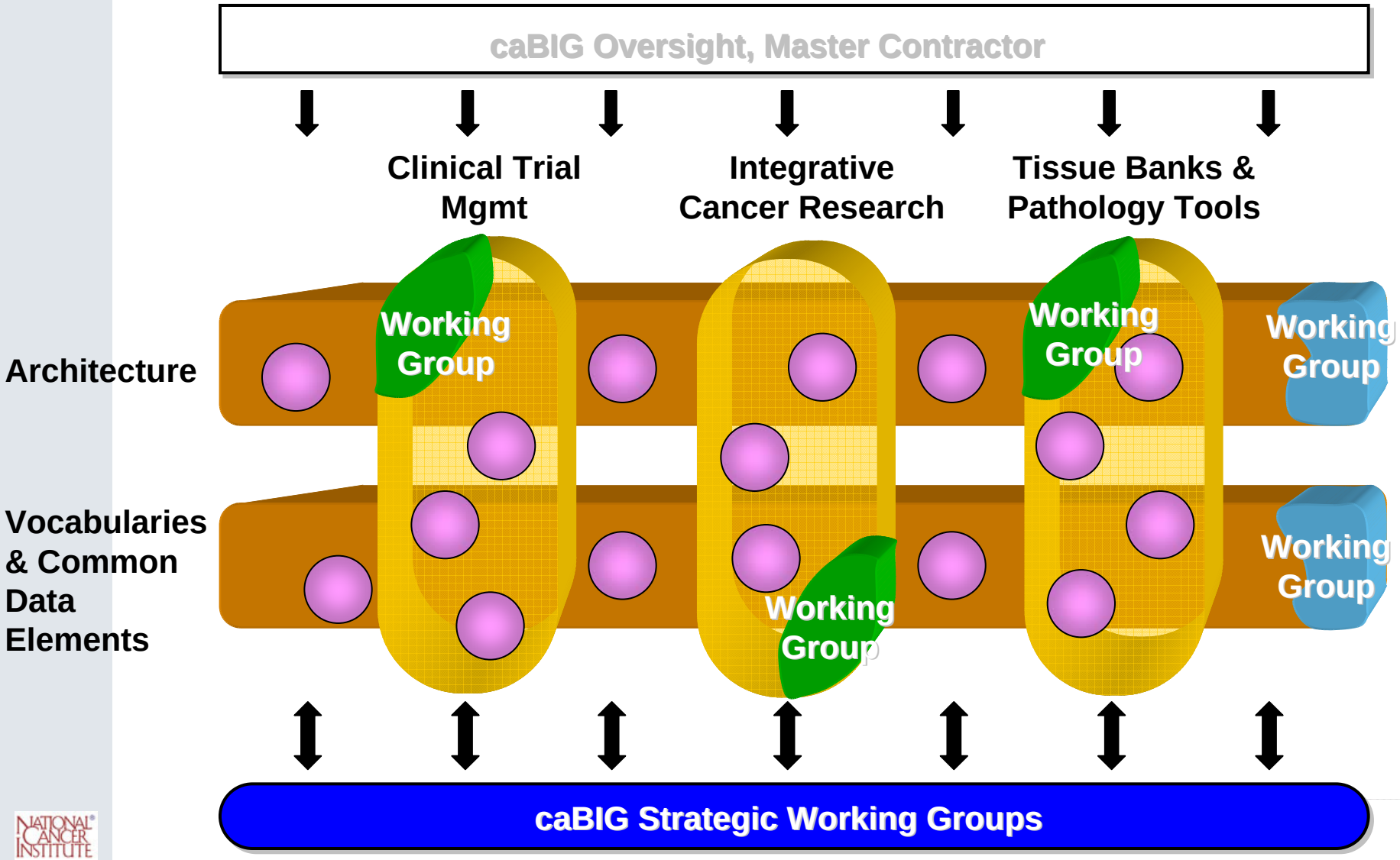
- The fewer, the better
- Let's not spend all of our time writing and re-writing parsers
- Must support CDE associations in order to convey all necessary semantic content and accompanying metadata

Entrance ramps: Cross-cutting workspaces

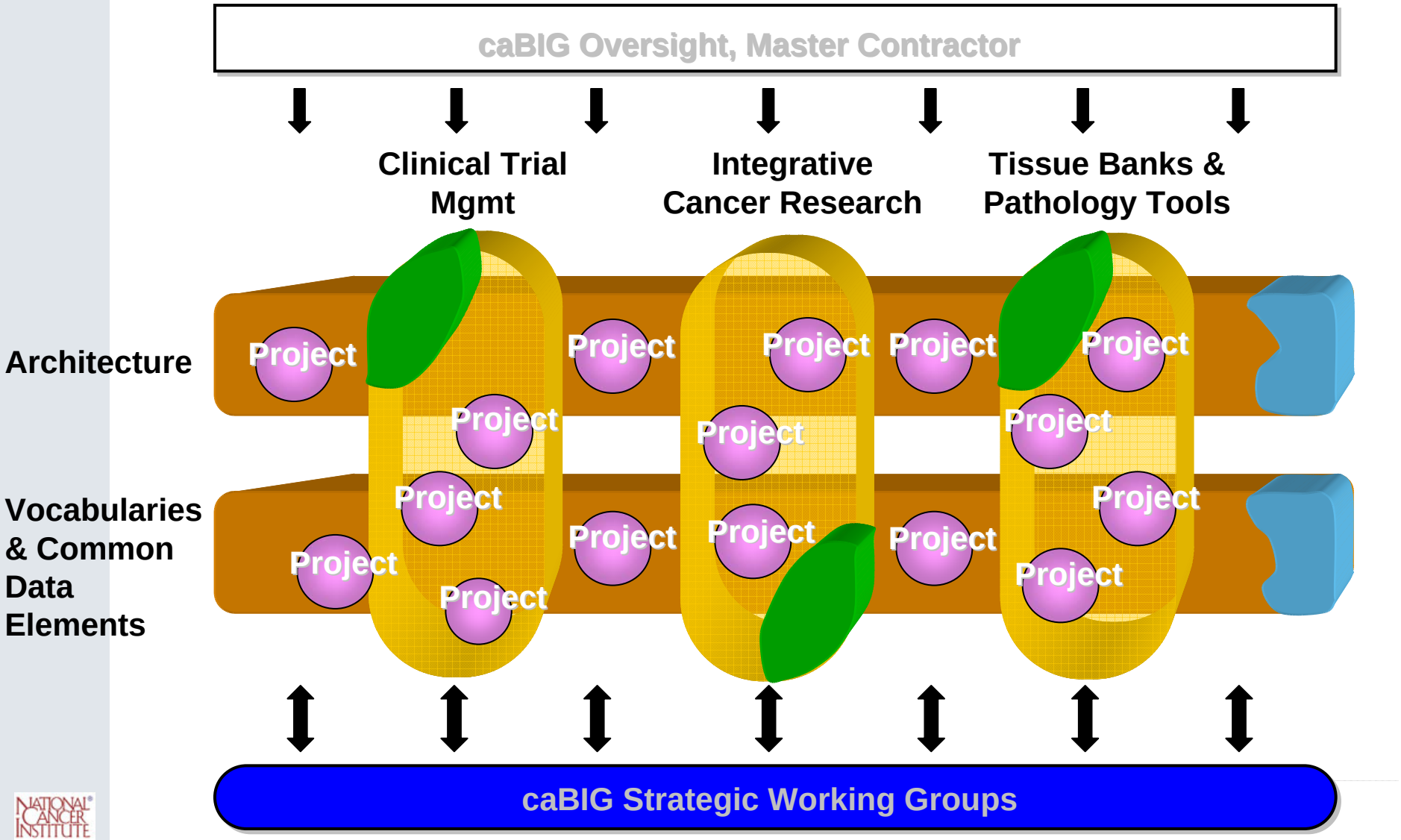
Cross-cutting Workspace Goals

- Identify, develop and publish common standards needed for semantic and syntactic interoperability
- Assist Domain workspaces with implementation of these standards
- Repeat

Urban planning



Construction, Manufacturing, Servicing, Consumption



Let’s Drive!


CHELSEA
CHELSEA PARK
GRAMERCY
GRAMERCY PARK
MADISON SQUARE PARK
BELLEVUE HOSPITAL
V.A. HOSP.
24TH STREET PARK
THEODORE ROOSEVELT BIRTHPLACE
STUYVESANT SQUARE
UNION SQUARE
UKRANIAN MUSEUM
ST. MARKS CHURCH
COOPER UNION LANDMARK
TOMPKINS SQUARE
EAST VILLAGE
EAST RIVER
EAST RIVER PARK
WEST VILLAGE
WASHINGTON SQUARE
NEW YORK UNIVERSITY
GREENWICH VILLAGE
WALKER PARK
NEW MUSEUM OF CONTEMPORARY ART
HAMILTON FISH PARK
SOHO
West Side Expwy
Franklin D. Roosevelt Dr
W. 29th St
W. 28th St
W. 27th St
W. 26th St
W. 25th St
W. 24th St
W. 23rd St
W. 22nd St
W. 21st St
W. 20th St
W. 19th St
W. 18th St
W. 17th St
W. 16th St
W. 15th St
W. 14th St
W. 13th St
Little W. 12th St
W. 12th St
W. 11th St
W. 10th St
W. 9th St
W. 8th St
W. 3rd St
W. Houston St
E. 29th St
E. 28th St
E. 27th St
E. 25th St
E. 24th St
E. 23rd St
E. 22nd St
E. 21st St
E. 20th St
E. 19th St
E. 18th St
E. 17th St
E. 16th St
E. 15th St
E. 14th St
E. 13th St
E. 12th St
E. 11th St
E. 10th St
E. 9th St
St. Marks Pl
E. 8th St
E. 7th St
W. 6th St
E. 6th St
E. 5th St
E. 4th St
E. 3rd St
E. 2nd St
E. 1st St
E. Houston St
11th Av
10th Av
9th Av
8th Av
7th Av
Av of the Americas (6th Av)
5th Av
Madison Av
Park Av South
Lexington Av
Irving Pl
3rd Av
2nd Av
1st Av
Levy Pl
Av A
Av B
Av C
Av D
Szold Pl
Broadway
University Pl
Greene St
Mercer St
Lafayette St
Bowery
4th Av
Astor Pl
Stuyvesant St
Bond St
Washington Sq N.
Washington Sq S.
Washington Pl
Waverly Pl
Bleecker St
MacDougal St
Sullivan St
Prince St
Gansevoort St
Horatio St
Jane St
W. 12th St
Bethune St
Bank St
W. 11th St
Perry St
Charles St
W. 10th St
Christopher St
Barrow St
Morton St
Leroy St
Clarkson St
W. Houston St
King St
Charlton St
Hudson St
Washington St
Greenwich St
Greenwich Av
Gay St
Grove St
Bedford St
Commerce St
Jones St
Cornelia St
Carmine St
Downing St
Patchin Pl
W. 4th St
Stanton St
Rivington St
Delancey St
Allen St
Orchard St
Ludlow St
Essex St
Norfolk St
Suffolk St
Clinton St
Attorney
Ridge St
Pitt St
Columbia St
Baruch Pl
Mangin Pl
Mott St
Elizabeth
58
57
54
53
52
51
49
48
46
45
42